Higher Education **46:** 37–59, 2003.

Different courses, different students, same results? An examination of differences in study progress of students in different courses

S. BEEKHOVEN*, U. DE JONG & H. VAN HOUT
SCO Kohnstamm Instituut, University of Amsterdam, Faculty of social and behavioral sciences, P.O. Box 94208, 1091 GE Amsterdam The Netherlands
*(*author for correspondence, e-mail: sandra@educ.uva.nl)*

Abstract. An important issue in higher education research is how to keep study progress at a good pace. In this article we will deal with the study progress of first-year students in various courses in Dutch higher education. Why are some courses more effective than others? Do such aspects as the composition of student population and different curricula influence variation in study progress, after controlling for individual factors? Multilevel analysis shows that there is in fact such variation between courses and this variation is only partially explained by individual characteristics and course characteristics. At the individual level, sex, initial ability, academic fit, expectation and commitment are important factors. After controlling for these individual factors, some courses still turn out to be more effective than others in getting their students to earn credits. Students in courses with a high proportion of women make more progress than students in courses with a high proportion of men. Furthermore, courses with a high average number of student study hours per week do better. Courses are ranked in an 'order of effectiveness' before and after important factors in the models are taken into account. Comparing courses before and after controlling for any characteristics clearly provides a different picture of effective courses.

Keywords: course characteristics, higher education, multilevel analysis, rational choice, school effectiveness research, student characteristics, student integration models, study progress

Introduction

An important issue in higher education research is how to keep study progress at a good pace. In this article we concentrate on the study progress of first-year students in various courses in higher education in the Netherlands. Our aim is to examine which factors at the individual level and at the level of study programs influence study progress. We are especially interested in the variation in progress between different courses. Do such aspects as different curricula, instruction and composition of student population influence study progress, after controlling for individual factors? We start by discussing some relevant theoretical concepts.

Theoretical background

School effectiveness research and higher education

School effectiveness research is an important field of research in several European countries, particularly in the United Kingdom (Teddlie and Reynolds 2000). However, this field of research does not include higher education. School effectiveness research, SER, concentrates on preschool, primary and secondary education and its purpose is to examine the net effect of schools on pupils. This is determined by comparing the output of schools, controlled for effects at the individual pupil level, such as socio-economic characteristics and prior abilities. 'Output of schools' usually refers to the knowledge and capacities pupils acquire between entering and leaving school. Other possible criteria for evaluating outcome are social skills or the level of educational attainment subsequently realized. A crucial indicator for the level of educational attainment is whether or not a person leaves secondary education with a diploma. In higher education the amount of credits is the most important and obvious outcome of education. In all courses in the Netherlands one credit represents a study load of 40 hours. Every higher education diploma represents the same amount of credits: 168 credits. When examining study progress of different courses, we look at the effectiveness of these courses and their study programs. In such a research, the data are hierarchical by nature and that is an important similarity to regular school effectiveness research. Pupils in primary and secondary education are nested in classes and classes are nested in schools. Students in higher education are nested in courses, and courses are nested in departments and universities. In view of this similarity we will use the analytical approach commonly used in school effectiveness research. This approach endeavors to distinguish the variance in outcome measures at each hierarchical level. The hierarchical level 'course' represents the context that individuals have in common; such as the building, the learning environment, the study culture etcetera. There are not only differences within the group of students (individual level), but also between that group and other groups (course level), for example between the sociology and psychology students. Therefore, we can split the total variance into two parts: variance at the individual level and variance at the course level.

Student integration models and rational choice

In most educational research on dropout and progress in higher education, theories are used that focus on the integration of students in higher education. This specific field of research is especially strong in the United States and the Netherlands. Student integration theories explain attrition as a lack

of agreement between standards and values of students and those of their study environments and therefore, as unsuccessful integration. Students have to interact with several actors involved and must try to feel at home. If students do not 'fit in', they are more likely to leave. Apart from integration, socio-demographic characteristics and subjective factors such as motivation and commitment are taken into account in these models. A lot of empirical research based on Tinto's 'model of institutional departure' (Tinto 1987) has been carried out. The construction of these models was repeatedly altered or supplemented with other variables (see for instance Pascarella et al. 1983). These attempts did not, however, always improve the explanatory power of the model.

In Amsterdam, a specific version of the student integration model has been tested several times (De Jong et al. 1997). In addition to attempting to adjust and improve the model, researchers were looking for other useful theoretical insights to explain differences in study careers (Need and De Jong 2001a). Beekhoven et al. (in press) combined the student integration model with aspects of the theoretical concept of rational choice theory. Rational choice theory states that individuals make choices based on a cost benefit analysis. This analysis is made within the social structure in which individuals operate. For students this process of rational choice means they have to estimate the costs and benefits of studying within their study environment in which they are integrated to a certain amount. Combining this rational choice concept with the integration concept in an interactive manner in a new model, explained more variance in academic progress than models based on each theoretical concept did separately (Beekhoven et al. in press). The most important variables from both concepts were the likelihood of success as perceived by students in the rational choice concept, and the 'academic fit' scale of the integration concept.

Study environments

Although many researchers hypothesize the effects of study environments on individual study outcomes, the evidence of the effects of study environments is not straightforward. A theoretical perspective on differences between departments or subjects can be found with Van Hout (1996), Biglan (1973 a, b) and Becher (1989). Biglan makes a distinction between cognitive aspects of subject matters in three dimensions, namely a hard/soft dimension, an applied/non-applied dimension and finally a life/non-life dimension. The last dimension, which distinguishes between life (for instance psychology and biology) and non-life (for instance chemistry and mathematics) subjects, is the most difficult one in making clear distinctions in subject matters. Becher has developed two dimensions to classify departments on a social aspect,

namely convergence/divergence and urban/rural. In both classifications the central idea is that the natural sciences share common research methods and have the same types of research goals, work fast (as life in an urban environment) and support one another. Anthropologists and historians as representatives of departments at the other end of the line, however, do not share research methods, have lots of space for choosing a research subject and do not need to rush to tell the world about their findings (as life in a rural environment).

However, the notion of an environment in which the student studies often only exists in theory. Most empirical studies analyze students as a single group without differentiating between courses or departments. The academic progress of students of different courses has rarely been compared. Although some researchers did focus on organizational characteristics that clearly belong to either courses or universities, they analyzed them as individual characteristics.

Explaining effects of organizational characteristics on the basis of an analysis performed at the individual level, through disaggregating the data, is known as the atomic fallacy (Hox 1994). Volkwein et al. (2000) treat organizational variables such as size and selectivity of institutes as individual characteristics. They perform a regression analysis, thus ignoring the fact that their respondents are nested in universities, which means that they ignore the fact that some students have more in common than other students. Van Den Berg and Hofman (2000) merged different courses into sectors and added them as dummy variables to their model together with other study program characteristics. It was concluded that students in natural sciences courses needed more time to finish their studies than students of other courses, even though the former group of students spent more hours studying. In the Netherlands, Prins (1997) paid attention to organizational factors using a method based on a comparison of effects per group, which meant he could only test one factor at a time. One study that did not disaggregate organizational factors, because it was based on a multilevel approach, showed that some study program characteristics influenced students' academic progress (Van Der Hulst and Jansen 2000). One of their findings was that the number of exams students have to pass had a negative effect on study progress.

Ultimately, the important question to be answered is which part of the variation in progress is caused by differences between individuals (within groups) and which part by differences between study programs (between groups). To establish the amount of variation at different levels, multilevel analysis has to be used. Need and De Jong (2001a) performed such an analysis. Their approach resembled school effectiveness research as applied to higher education. They had data on various departments in several univer-

sities. First, they determined the amount of variation at both levels if no other variables were accounted for. The variation at the study program level was small, only 5%. This variation was reduced when individual characteristics were taken into account. Their next step was to examine characteristics of study programs. Among some of the effects they found was, for example, the finding that students who made more progress were in study programs with more traditional[1] students and/or in study programs in which students were, on average, more satisfied with their education. After these characteristics were entered into the analyses, there was no variation left between departments. This has some interesting repercussions, as these results indicate that in the Netherlands it does not matter which department in which university one attends. When dealing with the same quality and composition of students, the universities do not differ in their effect on the study progress of individual students. This research was limited in that using the data of Need and De Jong afforded analysis only at a departmental level, not at a course level. In this article we want to look at differences in study progress between courses before and after controlling for individual characteristics.

Research questions:

1. How much of the variation in study progress is at the individual level and how much at the course level?
2. Is there variation in study progress at the course level after controlling for individual characteristics?
3. To what extent can this variation be explained by course characteristics?

To analyze these questions properly, we will use a technique called multilevel analysis, which allows us to distinguish in a statistically correct way between the individual level and the course level.

Hypotheses

The average variance between schools found in school effectiveness research in secondary education in the Netherlands ranges from 10 to 12 percent (Teddlie and Reynolds 2000). Empirical research (Need and De Jong 2001a) showed a variance of 5 percent[2] in academic progress in higher education at the departmental level. However, we believe that if the analysis is focused on a level closer to the individual, the course level, more variation will be found. Our first hypothesis is:

1. We expect to find more than 5% variation in study progress at the course level in higher education.

Individual characteristics

The conceptual model of combined integration and rational choice variables implies several chains of causal effects. Using a multilevel analysis, which is necessary to answer our specific questions, we are faced with the disadvantage that differentiating between direct, indirect and total effects can only be achieved by using very complex methods of analysis (Hox 1994). Furthermore, from other research (Beekhoven et al. in press) we know the indirect and total effects that can be expected in such an analysis of study progress. We therefore limit our analyses to direct effects. We expect the following direct effects on study progress of students in the Netherlands:

2. We expect to find that women progress through their studies faster than men. Numerous studies have confirmed the effect of sex on study progress (De Jong et al. 1997; Shah and Burk 1999). In most research the effect of sex on progress continues to exist even after controlling for various explanatory variables. This means that there are characteristics that should account for the advantage of women but these are not measured in the models.

3. High socio-economic status, SES, which is indicated by the income and educational level of parents, usually has no direct effects on educational achievement. SES can have indirect effects however, through subjective success chance for instance. We expect that no direct effect of SES will be found in this analysis.

4. We expect that students who have entered higher education through a traditional track will progress more rapidly through their studies. Nontraditional students are students with a deviant entry qualification. They are, in general, older than traditional students and are often found to make less progress (De Jong et al. 1997). However, direct negative effects of nontraditional students are not always found in the Netherlands (Need and De Jong 2001b).

5. We expect to find that students who repeated one or more classes in secondary education will progress more slowly. In the Dutch elementary and secondary school system it is not uncommon to let students repeat one or two classes in their school career. We consider this a variable indicating the initial ability of the student. Students who never repeated a class, have a higher initial ability and will do better in higher education (De Jong et al.1997).

6. Grade point average in secondary education[3] has a positive effect on study progress. Grade point average in secondary education also indicates initial ability (De Jong et al. 1997). (In the Netherlands, in secondary education, students can choose subjects from a nationally prescribed curriculum and each subject is partly tested in national tests.)

7. In the Netherlands, ethnic minority students are known to study at a slower pace, but they do not have a higher drop out rate than Dutch students

(Boogaard 1997). Speaking of ethnic minorities, it is important to note that they consist of different ethnic groups that cannot be treated as one homogenous group (Grayson 1998). Most common ethnic definitions use country of birth to categorize ethnic minorities. However, these definitions leave no room for mixed heritage or for an individual to express their own identity. An alternative option to avoid being categorized is to ask people their ethnic self-definition (Kinket and Verkuyten 1997). This means people who 'officially' (following the definition by birth) belong to an ethnic group, but consider themselves to be Dutch, can indicate this by choosing the category 'I feel Dutch'[4]. We expect students who identify themselves as belonging to an ethnic minority group to make less progress.

8. We expect to find a positive effect of students who feel more integrated. In Tinto's theory, the integration of students plays an important role. Students who feel more at home in their new environment are thought to be the better students and to make more study progress (Tinto 1987).

9. We expect to find a positive effect of students that are more committed to study progress. Students who give their study a central place in their lives and really try to work in a disciplined manner, will make more study progress.

10. We expect to find a positive effect of students who have higher expectations about their subjective chance of success on study progress. This proved to be one of the strongest predictors for actual study progress in studies by De Jong et al. (1997) and Beekhoven et al. (in press). Apparently, students are very well able to analyze their situation and their chances of achieving their goals.

Course composition variables

What do we expect of the factors at the course level? The following hypotheses will be tested:

11. Since women are known to study faster, we expect a high proportion of women in a course to have a positive effect on study progress. The assumption is that the features that account for the faster study pace of women have a positive effect on their male students.

12. Since the level of integration of individual students in a particular course is found to have a positive influence on study progress, we may find that a high mean level of integration in a course also contributes positively to individual study progress.

Course characteristics

13. We expect the number of scheduled hours for students to have a positive effect on integration at the individual level, jointly forming a cross-level effect

on study progress. A cross level effect means that a variable at the individual level (integration) interacts with a variable at the course level (number of scheduled hours). The idea is that students who are more integrated study faster and show a lower drop-out incidence, while the larger the number of scheduled hours is, the greater the opportunity will be for interacting with other students and, consequently, for becoming integrated in the study community.

14. Some research showed that in studies where students had to pass a large number of exams, the students were slowed down to a greater extent (Van Der Hulst and Jansen 2000). We expect a negative effect from this course level factor on study progress.

Data and methods

In 1998, a stratified sample was drawn from a cohort of first-year students of one university, 'Universiteit van Amsterdam', UvA, and two universities of professional education, 'Hogeschool van Amsterdam', HvA, and 'the Amsterdam school of business', HES[5]. The universities provided data on study progress of these students on an annual basis. From other research (De Jong et al. 1991) it is well known that students with poor results tend to respond less to questionnaires. To minimize the effects of this phenomenon, we attempted to over-sample the students with poor results in their first trimester. Further analyses showed that this procedure was slightly helpful. Overall, the students who responded earned more credits than other students (for details see Beekhoven, internal report 2001). The sample consisted of 1992 students.

In March 1999, the students were sent the first part of our survey, a questionnaire. About 40%, 782 students responded. All courses and the percentages of the response to the questionnaire are shown in Table 1. Unfortunately, the response is not evenly spread among the courses. For some courses the response is only 20%, while for others it is much better, about 50%. Some courses had too few students, so that for the purpose of our analysis we merged the university mathematics students with those of computer science, and those of economic science with those of financial economic science[6]. The first survey contained a lot of variables and scales based on concepts of both integration and rational choice theory. For integration we use the scale academic fit that relates to the match experienced between students and their course. The scale commitment is about the way in which students study and feel committed to their work. For rational choice, students reported their perceived likelihood of being successful in earning all 42 credits of the first-year program in one year. The variable is expressed as a

Table 1. Response percentages to the questionnaire

Courses	%	N
Amsterdam school of business: university of professional education		
finance and economics	32.4	18
marketing	26.0	38
business information studies	19.2	25
international business and languages	37.7	32
Hogeschool van Amsterdam: university of professional education		
nursing	60.0	45
cultural and social education	42.0	42
social legal services	42.7	32
laboratory science	42.7	32
computer sciences	21.3	16
University of Amsterdam:		
biology	52.5	21
physics and astronomy	39.5	17
mathematics	27.3	3
computer sciences	27.7	13
town and country planning	51.0	25
education and childhood studies	53.6	30
chemistry	45.5	10
science in society, propaedeutic year	51.7	31
medical biology	48.0	36
psychology	34.0	34
political science	38.7	29
law	37.0	37
economics	36.0	27
finance and economic science	28.1	9
human geography	56.0	42
communication studies	44.0	33
medicine	50.0	50
Dutch language and culture	40.0	26
history	38.6	29

percentage, ranging from 0 to 100. Table 2 shows all the variables and scales we use at the individual level.

The dependent variable is the number of credits students have earned after the first year. With 42 credits a student has passed the first-year exam, and will than receive the propaedeutic certificate.

The first questionnaire also invited students to participate in a qualitative part of the research. A fair number of students (50) accepted, and 25 of them subsequently participated in this qualitative research. The object of the qualitative research is to gain a good picture of the types of associations students make when dealing with the questions in the questionnaire. We cannot elaborate here on this part of our study, but we will quote some of their answers about expectations of success and academic fit. The first quotation is from a student in the marketing course answering the question: Can you tell us which factors your expectations of success are based on? "To tell you how I estimate my chances of success is easy, I think of the credits I have already earned, my motivation to continue with the course and I think of my self-confidence."

To give an example illustrating the concept of 'academic fit', we quote a student on the psychology course on one item in the 'academic fit' scale concerning the atmosphere on his course: "If today I had to rate the 'atmosphere on my course', I would give it the highest score. The atmosphere is really fantastic, because for a couple of months now I have been working on a research assignment. This research involves working with some other students. We are having a great time and we all are very enthusiastic. We have an excellent rapport with the teacher. If I have to evaluate the atmosphere at my course, I think of my relations with fellow students and with teachers."

Table 3 shows the course level variables. Firstly, some course composition variables that contain data on the proportion of women, ethnic minorities, nontraditional students, mean high school GPA and mean score on the academic fit scale. Secondly, characteristics of course programs; scheduled hours, number of exams and mean of hour's students of a course spent studying weekly (self reported in the survey). Most data about courses were collected from study guides and internal reports of the institutions.

The analyses were performed using Mlwin program version 1.02 (Goldstein et al. 1998). Because of several missing values on individual variables, 79 cases were removed. Furthermore, 53 students had filled out the survey about a course other than the one they had signed up for according to the data we received from the student administration. We started analysis with 650 students. Table 4 shows the mean of credits after the first year per course of the 650 students in our analyses.

The students included in the analyses differ from the students not included in the analyses. The former received more credits in the first trimester and

Table 2. Variables at the individual level

Sex	40% men	60% women
Grade point average in high school	mean 6.7	sd 0.6
Educational level of parents (range 1: lower education to 4: higher education)	mean 2.6	sd 1.1
Income parents (mean income of both parents in guilders)	mean 3793	sd 1118
Ethnic self-definition (group in category 'not Dutch' are students who felt they belong to an ethnic group or felt both Dutch and belonging to an ethnic group)	Dutch 87%	not D. 13%
Traditional student (standard entry qualification)	yes 83%	no 17%
Repeating one or more classes	no 71%	yes 29%
Expectation perceived likelihood of achieving 42 credits in one year (percentage from 0 to 100)	mean 73	sd 26.4
Credits earned in first trimester for each student we calculated the deviation from the mean of students in his or her course. This was necessary because in most but not all courses 14 credits could be attained in the first trimester. If a student earned more credits than average, he or she has a positive score.	mean .83	sd 43
Scales Items were scored on an 11-point scale (0 'this does absolutely not fit me or my situation' to 10 ' this fits me or my situation exactly'.) Scores of items with a ∗ were reversed.		
Academic fit	mean 7.3	sd 1.6
I feel at home here		
I do not like my fellow students∗		
I still feel awkward walking around the faculty∗		
I think I would fit in better with students of another course∗		
There is always a student who I can turn to with a problem		
I feel like one of the herd∗		
The atmosphere is good		
Cronbach alpha .80		
Commitment		
I find it hard to plan my work∗	mean 5.3	sd 1.7
Activities outside my study prevent me from studying∗		
I am satisfied with the commitment I had so far		
I find it hard to concentrate on subjects I find less interesting∗		
I tend to postpone study obligations∗		
I am a disciplined worker		
I can only concentrate in fits and starts∗		
My concentration is usually good		
I know I should spend more effort on studying but I can't seem to do it∗		
Study as fast as I can with the lowest effort, that is my motto∗		
Cronbach alpha .84		
Dependent variable: *Study progress; credits after one year of study*	mean 34.5	sd 10.9

sd = standard deviation.

Table 3. Course level variables

Course level composition variables
• *Proportion of women*∗: mean 54%, minimum 5%, maximum 91%
• *Proportion of ethnic minorities*∗ mean 14%, minimum 1%, maximum 44%
• *Proportion of nontraditional students*∗: mean 19%, minimum 0%, maximum 62%
• *Mean of secondary school grade point average*: mean 6.8, minimum 6.5, maximum 7.8
• *Mean of academic fit*: mean 7.3, minimum 6, maximum 8
Course program characteristics
• *Level* (universities of professional education (value 0) and universities (value 1), the former is of a lower educational level): The data contains 9 courses of universities of professional education and 15 courses of university
• *Percentage of scheduled hours*: mean 33%, minimum 11%, maximum 55%
• *Number of exams*: mean 12, minimum 3, maximum 31
• *Mean of hours students spent studying*: mean 28.6, minimum 16, maximum 40

∗ based as much as possible on population data, otherwise on our survey data.

after the first year. T-test analyses showed that this difference is significant. A chi-square test showed that they also differ significantly in the percentage of traditional students; the group not included contains a higher percentage of nontraditional students, 31% as compared to 17%. The two groups do not differ in their distribution over courses. So the analyses are biased towards the somewhat better students, but this bias is evenly distributed over all courses. The fact that data are missing on poorly performing students is not uncommon in study progress research. De Jong et al. (1991) showed that in their study using a file with weighed cases to compensate for this aspect did not effect the strength of the parameters in the models.

Results

Percentages of course variation

We will first test Hypothesis 1, about the amount of variation at the course level and we therefore start the analysis with a so-called 'intercept-only model'. This means that Model A contains no other variables than the dependent variable. The results are presented in Table 5. For the 'intercept only model', Model A, only an intercept (or constant) and the two variance components are shown. As can be learned from the variation components of Model A, there is a nearly 14% variance in study progress at the course level. This is the percentage of total variance in study progress that can be attrib-

Table 4. Credits per course after one year (university courses in *italics*)

Mean of credits after one year, ranked from high to low, maximum = 42	mean	sd	N
cultural and social education	39.85	6.00	40
nursing	39.57	5.65	37
marketing	38.69	7.93	31
biology	38.29	6.12	17
science in society, propaedeutic year	37.52	5.98	25
education and childhood studies	37.50	6.76	28
business information studies	36.88	6.19	19
town and country planning	36.73	10.02	22
laboratory science	36.45	7.25	29
medicine	36.09	10.02	47
international business and languages	35.88	8.75	29
Dutch language and culture	35.29	10.00	24
psychology	34.89	10.42	28
finance and economics	34.74	6.31	16
medical biology	34.46	8.84	26
social legal services	34.28	10.79	29
chemistry	33.80	8.32	10
human geography	33.79	9.76	36
physics and astronomy	31.23	13.26	11
communication studies	30.00	13.45	28
history	29.83	12.84	21
computer sciences	28.33	17.39	12
law	26.90	16.66	31
political science	25.67	11.78	21
economics	23.90	15.95	21
mathematics and computer sciences	22.83	15.99	12
total	34.23	10.98	650

sd is standard deviation, N is number of students.

uted to variation between courses[7]. The other 86% is the variation between students. Clearly, there is enough variance at the course level to continue the analysis.

Individual factors

In Model B we included variables that were measured at the individual level. In Table 5 the results of Model B are divided into a fixed part, comprised of the intercept, unstandardized regression coefficients and their standard errors,

Table 5. Results of multilevel analyses, Model A through D. Standard deviation between brackets

	Model A	Model B	Model C	Model D
fixed effects				
intercept	33.77	20.58	19.95	18.66
individual variables				
sex		1.76	1.48	1.49
		(0.67)	(0.69)	(0.69)
academic fit		0.48	0.49	0.49
		(0.19)	(0.19)	(0.19)
commitment		0.51	0.50	0.48
		(0.18)	(0.18)	(0.18)
subjective chance propaedeutic certificate		0.09	0.09	0.09
		(0.01)	(0.01)	(0.01)
repeating class(es)		–2.00	–2.10	–2.01
		(0.65)	(0.64)	(0.65)
no Dutch identity		–2.50	–2.54	–2.56
		(0.91)	(0.91)	(0.91)
credits in first trimester		1.11	1.10	1.12
		(0.08)	(0.08)	(0.08)
course level variables				
level			–3.13	–1.82
			(1.40)	(1.33)
percentage women			0.06	0.07
			(0.03)	(0.03)
average amount of hours studied				0.26
				(0.10)
variance components				
level 2: courses	16.59	13.62	8.7	6.36
level 1: students	104.93	50.98	50.59	50.56
variance at level 2	13.60%	21.20%	14.60%	11.20%
explained at level 1∗		51%	51%	51%
explained at level 2∗		18%	48%	61%
deviance	4909.69	4446.91	4437.23	4430.635
df		7	2	1
significance		< 0.00	< 0.01	$<.001$

∗compared to Model A.

and a random part, comprised of variance components at the individual and course level. To limit the number of variables, we removed variables that had no significant effects, as recommended by Hox (1994). The significant effects confirm our hypotheses about the positive effect on study progress of being a woman, the negative effect of repeating classes, the positive effect of being more integrated, committed and having higher expectations of success. The hypothesis about the influence of parental education and income was also confirmed; the direct effects proved not significant. This does not mean that SES has no effect, but positive and negative effects may have neutralized any such effect.

We want to elaborate on the effect of ethnic self-definition. We could not separate students who feel partly Dutch and partly that they belong to an ethnic group from students who feel they belong only to an ethnic group, because then the two latter groups would be to small to give reliable results (Tabachnick and Fidell 1996). Students who did not feel Dutch had 2.5 fewer credits than other students. If the more commonly used variable dividing people into ethnic minorities on the basis of the country of birth of (one of) their parents is used, we find no significant effect. Self-definition is therefore more important in explaining differences in study progress than the official definition. The group of students who feel Dutch contains students who are not Dutch by official standards. It is important to note that working with standard definitions can give a distorted view of this complex aspect of peoples' lives (see also Beekhoven et al. 2000).

The seven significant variables in Model B reduced the variance between students significantly and explained 51% of the variation at the individual level. The percentage of variance at the course level increased to 21%, which means that after taking into account the individual characteristics of students, an even larger percentage of the variance left is due to relative differences between courses.

Course composition

In Model C we added the course composition variables and the variable level (type of higher education). The well-known fact that students in universities of professional education (where participating in classes is partly compulsory) make more progress than university students was confirmed by our analyses. Only two of the five course composition variables in Table 3 proved to have any significant effects. There is a significant positive effect of the percentage of women in a course. The effect of the high proportion of women confirms our 11th Hypothesis. The effect co-exists with the individual effect of sex. The other hypothesis: a positive effect of high mean score on academic fit in courses is not confirmed since this effect was not significant.

We have also checked if subject area had any effect as a course level variable. To accomplish this, we made dummy variables, representing the subject areas in social, economic, natural sciences, health-care and artistic sectors. The effects of these dummy variables were not significant. Apparently, differences in sciences such as described by Becher (1989) and Biglan (1973a) cannot be reflected in differences in study progress between sectors. The coefficients of the individual variables in Model C remain the same as in Model B. The variance at the course level has decreased to 14.6%. Including course composition variables explained 48% of the variance between courses compared to Model A.

Course characteristics

In our final Model D we explored possible influences of the course characteristics: percentage of scheduled hours and number of exams. We then added an extra characteristic: the self reported average number of hours students spend on course study per week (including visiting classes and workgroups). There have always been explicit common-sense ideas about the differences in student investment between courses, for example students in medicine have to invest many more hours than students in economics in order to graduate. We entered the average number of hours per week that students studied per course as an indicator of the 'work ethic'.

Neither the percentage of scheduled hours nor the number of exams had any significant effect and are therefore not shown in Table 5. The construction of the variable 'percentage of scheduled hours per week' led to some problems. The study programs at the universities of professional education involved in this study, all work with problem-based learning. This implies that students work a great deal in small groups and meet with a teacher on a regular basis. At the UvA, most courses work with lectures and/or working groups. It is not easy, however, to compare the educational systems used at the three universities in our data. This makes the variable 'percentage of scheduled hours per week' less valid and less reliable. Because it was not always clear whether an exam was a written test or not, the same can be said about the variable 'number of written exams' students have to pass.

The average number of hours per week students studied per course proved the only course characteristic that had any significant effect. Students in courses with a high average number of hours studied made more study progress. Similar to the way the proportion of women has a positive influence, being in a study environment where the average student works hard will have a positive effect on all the students' progress.

To test our hypothesis about interaction between 'percentage of scheduled hours per week' and 'academic fit', we calculated a new variable by

multiplying both variables. However, this variable had no significant effect and we had to reject Hypothesis number 13. We present the results of the model from which the non-significant course characteristics were removed. We now have Model D, which is significantly better than Model C, with an 11.2% variance left at the course level.

Different effects per course?

Multilevel analyses make it possible to check if the effect, or slope, of an individual characteristic is different for each course. Although we had no theoretical basis to suspect such differences, we wanted to check if the assumption of no differences in effects per course was correct. Therefore, the fixed effects (the coefficients or slopes) were made random, meaning they were allowed to differ per course. For example, it could be that the effect of sex was not equally strong in all courses. We checked if any of the seven slopes were significantly random, but this was not the case. This is a satisfactory result because it would be difficult to find explanations if for example, 'being committed' hardly matters in some courses and its contribution to progress would be large in other courses.

Predictions

We performed the analysis in four steps (Model A, B, C and D). After each step the predictions for the average amount of credits after one year of each course change, because they are controlled for the added variables. In Table 4, the courses were ranked from the highest average amount of credits to the lowest. This ranking can be read as the ranking of Model A, simply comparing the average study progress without controlling for any variables (the intercept only model). We can create this ranking also for the Models B, C and D. Table 6 shows the rankings as derived from each model. Looking at the ranking for Model B, in which individual factors are taken into account, the ranking of the courses changes, for some courses more dramatically than for others. Other changes occur when computing new rankings after entering the course level variables in Models C and D, respectively. For instance, psychology starts out at Position 13 in the first ranking. In Model B it shifts to Position 10, in Model C to 8 and in Model D psychology ends up in the third position. Furthermore, as can be seen for nursing, courses with a lot of women do well, but less so if the individual characteristic sex and the course characteristic percentage of women are taken into account.

The most important aspect of looking at the data in this manner is that we can point out that it would be unfair to compare courses without taking into account their student populations and a number of course composition

Table 6. Ranking (position 1 to 26) of courses on average amount of credits, per model. In each model the ranking is controlled for the different variables in the model.

Ranking in	Model A	Model B	Model C	Model D
cultural and social education	1	2	13	11
nursing	2	3	14	8
science in society, propaedeutic year	3	8	3	6
education and childhood studies	4	7	10	5
marketing	5	1	2	2
biology	6	9	4	9
business information studies	7	4	5	4
town and country planning	8	5	1	1
laboratory science	9	6	15	20
medicine	10	11	7	12
international business and languages	11	12	21	25
Dutch language and culture	12	17	12	10
psychology	13	10	8	3
finance and economics	14	14	17	14
medical biology	15	15	9	16
social legal services	16	13	18	21
chemistry	17	16	6	13
human geography	18	18	11	7
physics and astronomy	19	20	16	17
communication studies	20	19	19	15
history	21	21	20	18
computer sciences	22	22	22	19
law	23	23	23	24
political science	24	24	24	23
economics	25	26	26	25
mathematics and computer sciences	26	25	25	26

variables. If a course does well because of its population, it is not really an accomplishment of their study program.

Good and bad practices

Now that we have established that there are differences in study progress of students of different courses, what is the extent of differences between courses? To find out, we make a plot of the residuals of each course compared

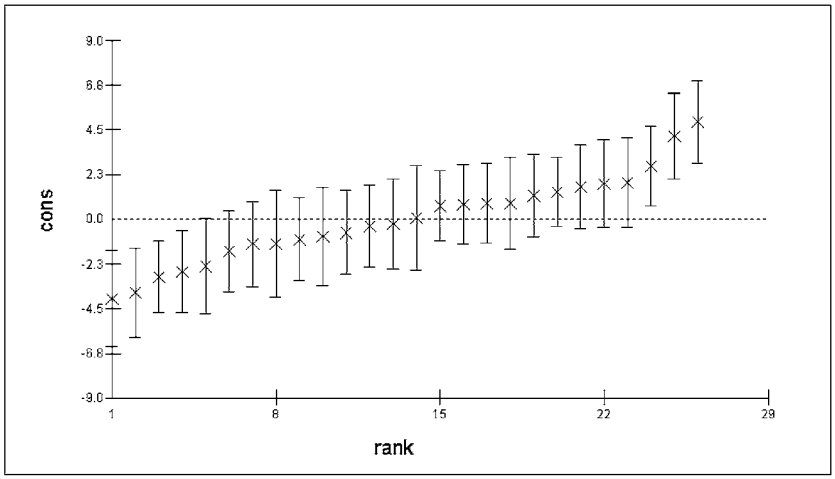

Figure 1. Plot of residuals of Model D
The three courses represented by the error bars in the left of picture, which have the lowest average credits are: (26, 25, 24) mathematics and computer sciences, international business and languages and law.
The three courses represented by the error bars in the right of picture, which have the highest average credits, are: (3, 2, 1) psychology, marketing and town and country planning.

to the grand mean (the average of the whole group). This will provide insight into the relative differences between courses. We will use the most complete model; Model D. As can be seen in Figure 1, the estimated residuals in Model D for the courses are below, equal or above the average estimate of the whole group. The stars in Figure 1 mark the estimated residuals per course; the lines crossing the stars are the error bars of standard deviations.

This method of studying residuals is common in school effectiveness research because it identifies schools that score under or above the average of all schools in the sample. Such results are often used to describe good and bad practices. If a particular course has an error bar that does not overlap any other error bar (Goldstein et al. 1998), the two course residuals are significantly different.

In our figure, the course that has the largest 'negative' distance compared to the average differs significantly from the 8 'best' courses. The 'best' course differs significantly from the 13 courses at the bottom of the ranking. It would be fair to consider the first course and the last three courses as examples of 'good' and 'bad' courses.

Conclusion and discussion

In this article we dealt with the study progress of first-year students in various courses in Dutch higher education. Leaning on the approach commonly used in primary and secondary education in school effectiveness research, we studied the effectiveness of different courses in getting their students to earn credits in the first year. To get a clear and methodologically solid picture of the amount of variation in study progress between different courses in relation to the variation in study progress between individual students we performed a multilevel analysis. As expected, in the first model that was analyzed we found a substantial amount of variance between courses in study progress after one year. Controlling for individual factors in the second model, the percentage of variance that could be attributed to courses was not reduced; on the contrary, this percentage of variance at the course level increased. Taking course level factors into account in the third model did reduce this variance between courses. However, after analyzing our complete model, over 10% of the variance in study progress remained at the course level.

A number of our hypotheses were confirmed. The individual characteristics had effects similar to those found in numerous previous studies. Women make more progress, students who repeated (a) class(es) and students without a Dutch ethnic self-definition make less progress. More committed students and students with a higher subjective chance of success make more progress. And, of course, those who started out performing better than the average student in the first trimester benefit from such a good start and earn more credits in the first year.

We could confirm only one of our hypotheses about influences at the course level: the hypothesis about the proportion of women in a course having a positive effect on progress of students of that course. This effect was often found in higher education studies. Some people are inclined to subscribe the advantage for women to their more serious nature and the fact that they spent more hours studying. However in this study there is no significant difference between the amount of hours men and women study weekly. One possible explanation might be that women study more effectively than men. Or men exaggerate the amount of hours they study. Another effect of a course level variable we found was that the average number of hour's students in a course study weekly has a positive effect on the number of credits the students earned.

Would the amount of variation between and within courses and the factors contributing to the variation be similar in other years of the study? We would like to reproduce this analysis with the number of credits students earned after their second year, or after subsequent years of study.

Now that a substantial amount of variance at the course level has been established in this study, it is important to reproduce these findings for courses at other universities and for other student cohorts. What should be borne in mind is that the analyses we performed relate to one cohort of students. Universities are constantly working on improving their study programs and renewing the content of their curriculum. Analyzing different cohorts could lead to different rankings. That way, the comment of a teacher: "This was a very bad year" could be put to the test.

In the final model there was still variation at the course level unexplained. What could be causing the almost 10% variation in study progress after one year that remains between courses in the findings we presented here? An obvious conclusion is that it is caused by differences in the way students are stimulated to adopt effective study behavior. In this article the method of school effectiveness research was used to explain difference in the important outcome factor of higher education; earning credits. Are all courses effective in the way the students' progress through the curriculum? No, some courses are more effective in getting their students to earn credits than other courses. This could be caused by less competent teachers or by a curriculum that is not stimulating for students. It could very well be possible to find other factors that can explain the variance at the course level if there were more variables about those courses in the analyses. For instance, the style of education, if and how much students are counseled, and so on. Jansen (1996) for example found effects in the way exams are chronologically spread over the year, the way subjects were programmed, (whether or not in parallel), and the number of re-examinations. It would take more information and more time to develop additional operationalizations of course characteristics, but it may be worth it.

We would like to conclude by emphasizing that in future research analyses should definitely be performed using a multilevel technique and using data clustering the individuals at the course level and not at some higher level. This gives a more realistic picture than using departmental level data.

Acknowledgments

We would like to thank Dr. Ariana Need of the University of Nijmegen for her useful comments.

Notes

1. Traditional students are students who followed a straight path to higher education. Non-traditional students are those who followed a different and therefore longer path because they did not have the standard qualification giving access to the course they wanted to participate in.
2. For universities of professional education the variation was 5%, for other universities 2.5%.
3. The grade point average is calculated from all the subjects (varying from 5 or 8) that the respondents reported their marks on, if there were at least 5 marks.
4. The categories students could choose from in the questionnaire were; a. I feel Dutch, b. I feel both Dutch and belonging to an ethnic group, namely . . . , c. I feel belonging to an ethnic group, namely . . . .
5. The Netherlands has two types of higher education: universities and universities of professional education, referred to as 'hogescholen'. The latter are more practically oriented and are of a lower level than universities, although for both education is based on 4 years. Universities of professional education can be compared to new universities in Britain.
6. The curriculum in the first year of these courses is identical.
7. For Model A: (16.59 / 104.93 + 16.59) $*$100

References

Becher, T. (1989). *Academic Tribes and Territories. Intellectual Enquiry and the Cultures of Disciplines*. Society for Research into Higher Education & Open University Press.

Beekhoven, S. (2001). *Method, Sampling, Response and Non-response.* Internal report SCO Kohnstamm Institute: Amsterdam.

Beekhoven, S., De Jong, U., Van Hout, H. (in press). 'Explaining Academic Progress via Combining Concepts of Integration Theory and Rational Choice Theory', *Research in Higher Education.*

Beekhoven, S., De. Jong, U. and Van Hout, H. (2000). 'Ethnic Identity, Integration and Study Progress'. *Presented at the Sociaal Wetenschappelijke Studie Dagen.* Amsterdam.

Biglan, A. (1973a). 'The Characteristics of Subject Matter in Different Academic Areas', *Journal of Applied Psychology* 57(3), 195–203.

Biglan, A. (1973b). 'Relationships Between the Subject Matter Characteristics and the Structure and Output of University Departments', *Journal of Applied Psychology* 57(3), 204–213.

Boogaard, M. (1997). *Van Buiten, Geleerd: Allochtone En Buitenlandse Studenten in Het Nederlandse Hoger Onderwijs* [From outside, Studied: Immigrants and Foreign Students in Dutch Higher Education]. Amsterdam: Het Spinhuis.

De Jong, U., Koopman, P. and Roeleveld, J. (1991). S*nelwegen en Slingerpaden in en om het Hoger Onderwijs.* [Highways and Side Tracks in Higher Education]. Report, Ministerie van Onderwijs Cultuur en Wetenschappen, The Hague, The Netherlands.

De Jong, U., Roeleveld, J., Webbink, H.D. and Verbeek, A.E. (1997). *Het Amsterdamse Studieloopbaanmodel.* Deel 6 Panelstudie onder Scholieren en Studenten Verder Studeren [The Amsterdam Model of Study Careers]. Report, Ministerie van Onderwijs Cultuur en Wetenschappen, The Hague, The Netherlands.

Goldstein, H., Rasbach, J., Plewis, D., Draper D., Brown, W., Yang, M., Woodhouse, G. and Healy, H. (1998). *A User's Guide to MLwin*. Bath: University of Bath.

Grayson, J.P. (1998). 'Racial Origin and Student Retention in a Canadian University', *Higher Education* 36, 323–352.

Hox, J.J. (1994). *Applied Multilevel Analysis*. Amsterdam: TT Publicaties.

Jansen, E.P.W.A. (1996). *Curriculumorganisatie En Studievoortgang* [Curriculum Organization and Study Progress]. Doctoral dissertation, University of Groningen, The Netherlands.

Kinket, B. and Verkuyten, M. (1997). 'Levels of Ethnic Self-Identification and Social Context', *Social Psychology Quarterly* 60(4), 338–354.

Need, A. and De Jong, U (2001a). 'Do Local Study Environments Matter? A Multilevel Analysis of the Educational Careers of First-year University Students', *Higher Education in Europe* 26(2), 263–278.

Need, A. and De Jong, U. (2001b). 'Educational Differentials in the Netherlands. Testing Rational Action Theory', *Rationality and Society* 13(1), 71–98.

Pascarella, E.T., Duby P.B. and Iverson, B.K. (1983). 'A Test and Reconceptualization of a Theoretical Model of College Withdrawal in a Commuter Institution Setting', *Sociology of Education* 56, 88–100.

Prins, J. (1997). *Studie-uitval in Het Wetenschappeljk Onderwijs. Studentkenmerken En Opleidingskenmerken Als Verklaring Voor Studie-uitval* [Drop out in University Education]. Doctoral dissertation. Nijmegen, The Netherlands: University Press Nijmegen.

Shah, C. and Burke, G. (1999). 'An Undergraduate Student Flow Model: Australian Higher Education', *Higher Education* 37, 359–375.

Tabachnick, B.G. and Fidell, L.S. (1996). *Using Multivariate Statistics*. California: Harper Collins College Publishers.

Teddlie, C. and Reynolds, D. (2000). *The International Handbook of School Effectiveness Research*. London: Falmer Press.

Tinto, V. (1987). *Leaving College. Rethinking the Causes and Cures of Student Attrition*. Chicago: The University of Chicago Press.

Van Den Berg, M. and Hofman, A. (2000). 'Studievoortgang in Het Wetenschappelijk Onderwijs' [Study Progress in Higher Education], *Tijdschrift voor Onderwijs Research.* 24(3/4), 288–307.

Van Hout, H. (1996). Verschillen Tussen Wetenschapsgebieden: Een Structureel Kenmerk van de Universiteit [Differences Between Subject Area's: a Structural Feature of Universities]. Amsterdam: Universiteit van Amsterdam.

Van der Hulst, M. and Jansen, E.P.W.A. (2000). 'Effecten van Curriculumkenmerken op Studievoortgang en Uitval van Studenten Techniek' [The Effects of Curriculum on Progress and Drop-out of Applied Science Students], *Tijdschrift Voor Onderwijs Research* 24(3/4), 237–247.

Volkwein, J.F., Valle, S., Parmley, K. and Blose, G. (2000). 'A Multi-Campus Study of Academic Performance and Cognitive Growth Among Native Freshman, Two-Year Transfers, and Four-Year Transfers'. *Presented at the Association for Institutional Research's Annual Forum,* Cincinnati, US.